Session C: Covariance (error correlations and their effect on error propagation)

C.1 Introduction

In statistics courses, and in the laboratory notes on error propagation, you will have come across the general formula for error propagation when y is a function x_1 and x_2 given below:

$$\sigma_y^2 = \left(\frac{\partial y}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial y}{\partial x_2}\right)^2 \sigma_{x_2}^2 \tag{C.1}$$

Exercise C.1

Verify that the formula C.1 leads to the following expressions for the error on y:

Function	Variance
$y = x_1 + x_2$	$\sigma_y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2$
$y = x_1 \times x_2$	$(\sigma_y/y)^2 = (\sigma_{x_1}/x_1)^2 + (\sigma_{x_2}/x_2)^2$
$y = x_1/x_2$	$(\sigma_y/y)^2 = (\sigma_{x_1}/x_1)^2 + (\sigma_{x_2}/x_2)^2$
$y = \ln(x_1)$	$\sigma_y^2 = (\sigma_{x_1}/x_1)^2$

This formula applies when x_1 and x_2 are independent. The problem is that x_1 and x_2 are not always independent. To take account of this we must extend our formula by including an additional term:

$$\sigma_y^2 = \left(\frac{\partial y}{\partial x_1}\right)^2 \sigma_{x_1}^2 + 2\left(\frac{\partial y}{\partial x_1}\right)\left(\frac{\partial y}{\partial x_2}\right)\mathrm{cov}(x_1, x_2) + \left(\frac{\partial y}{\partial x_2}\right)^2 \sigma_{x_2}^2 \tag{C.2}$$

The additional term involves the **covariance**, which we denote by $\mathrm{cov}(x_1, x_2)$. In the following we shall consider the variances and covariances of the *mean values* of x_1 and x_2, but the equivalent forms for the variances and covariances of their *populations* can be derived by analogy.

You should already know that the *variance* of the means of x_1 and x_2 are defined by:

$$\sigma_{x_1}^2 = \sum_{i=1}^{n} \frac{(x_1(i) - \overline{x_1})^2}{n(n-1)} \tag{C.3}$$

$$\sigma_{x_2}^2 = \sum_{i=1}^{n} \frac{(x_2(i) - \overline{x_2})^2}{n(n-1)} \tag{C.4}$$

The *covariance* of the means of x_1 and x_2 is given by:

$$\mathrm{cov}(x_1, x_2) = \sum_{i=1}^{n} \frac{(x_1(i) - \overline{x_1})\,(x_2(i) - \overline{x_2})}{n(n-1)} \tag{C.5}$$

If when we take our set of readings $x_1(i)$ and $x_2(i)$ there is no relationship between x_1 and x_2 then the covariance has an expectation value of zero and we say that x_1 and x_2 are independent, and we can use equation C.1. If there *is* a relationship between x_1 and x_2 then the covariance will be non-zero and we must use equation C.2. In this case we then say x_1 and x_2 are *correlated*.

C.2 Example

To reinforce this point we'll look at a straightforward example. Let's assume that the y-values of our data are given by $y = x_1 + x_2$, but where our variables x_1 and x_2 are trivially related by $x_2 = 2x_1$. Of course, you can see right away that we could equally well write $y = 3x_1$. It is obvious that x_1 and x_2 are dependent. Now, let's say that we perform some experiment whereby we obtain the following data:

Reading	1	2	3	4	5	Mean
x_1	10	9	11	12	8	10
x_2	20	18	22	24	16	20

Now we can calculate the variance of x_1 using equation C.3:

$$\sigma_{x_1}^2 = \frac{0 + 1^2 + 1^2 + 2^2 + 2^2}{5 \times 4} = 0.5$$

Similarly, the variance of x_2, using equation C.4, is:

$$\sigma_{x_2}^2 = \frac{0 + 2^2 + 2^2 + 4^2 + 4^2}{5 \times 4} = 2.0$$

We can also calculate the covariance term using equation C.5:

$$\text{cov}(x_1, x_2) = \frac{0 \times 0 + 1 \times 2 + 1 \times 2 + 2 \times 4 + 2 \times 4}{5 \times 4} = 1.0$$

Let us now calculate the expected variance of y. If we use (wrongly) equation C.1 we find:

$$\sigma_y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 = 0.5 + 2.0 = 2.5$$

Using equation C.2 we find:

$$\sigma_y^2 = \sigma_{x_1}^2 + 2\text{cov}(x_1, x_2) + \sigma_{x_2}^2 = 0.5 + 2 \times 1.0 + 2.0 = 4.5$$

In this case, we demonstrate that using equation C.2 is the correct approach, since we know that $y = 3x_1$. If we simplify our function and then calculate the variance of y (using equation C.1) we find:

$$\sigma_y^2 = 9 \times \sigma_{x_1}^2 = 4.5$$

As you might expect, this trivial example clearly shows that when parameters are dependent, we must take into account the covariance term.

C.3 The covariance matrix

In the case that our function is dependent on two variables, $y = f(x_1, x_2)$, we can write equation C.2 more succinctly in matrix form:

$$V_y = D^T V_x D \tag{C.6}$$

Here, V_x is the "covariance matrix" for the xs, and is given by:

$$V_x = \begin{pmatrix} \sigma_{x_1}^2 & \text{cov}(x_1, x_2) \\ \text{cov}(x_1, x_2) & \sigma_{x_2}^2 \end{pmatrix} \tag{C.7}$$

Note that the covariance matrix is symmetric.

D is the column vector $\begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \end{pmatrix}$ and D^T is its transpose.

Clearly the V_x matrix can be any size to accommodate any number of variables; $y = f(x_1, x_2, \ldots x_n)$. It is also worth noting that equation C.6 can also be used when several functions (i.e. $y_1, y_2, \ldots y_m$) are formed from $x_1 \ldots x_n$. In this case D is not a column vector, but an $m \times n$ matrix.

Let's take a look at an example of this in the exercise below:

Exercise C.2

1. Say we have two functions y_1 and y_2 that are dependent on the variables x_1 and x_2:

   $$y_1 = \frac{x_1}{(x_1 + x_2)} \qquad \text{and} \qquad y_2 = \frac{x_2}{(x_1 + x_2)}$$

   Let us take some numerical values, say $x_1 = 4$ and $x_2 = 6$ and let $V_x = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

   By choosing this value of V_x we declare that the uncertainty on x_1 and x_2 is one, and that x_1 and x_2 are independent, since the covariant (off-diagonal) term is zero.

   Show that:

   $$D = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} \frac{(x_1 + x_2) - x_1}{(x_1 + x_2)^2} & \frac{-x_2}{(x_1 + x_2)^2} \\ \frac{-x_1}{(x_1 + x_2)^2} & \frac{(x_1 + x_2) - x_2}{(x_1 + x_2)^2} \end{pmatrix}$$

   Note that, unlike V_x, D is not symmetric.

   Also show that if we put in the values for x_1 and x_2, $D = \begin{pmatrix} \frac{6}{100} & \frac{-6}{100} \\ \frac{-4}{100} & \frac{4}{100} \end{pmatrix}$.

2. Now determine V_y:

   $$V_y = \begin{pmatrix} \sigma_{y_1}^2 & \text{cov}(y_1, y_2) \\ \text{cov}(y_1, y_2) & \sigma_{y_2}^2 \end{pmatrix} = \begin{pmatrix} 0.06 & -0.04 \\ -0.06 & 0.04 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0.06 & -0.06 \\ -0.04 & 0.04 \end{pmatrix}$$

3. Finally, convince yourself that $y_1 = 0.4 \pm 0.072$ and $y_2 = 0.6 \pm 0.072$ and that the covariance term, $\text{cov}(y_1, y_2) = -0.0052$.

Notice that the matrix V_y is also a covariance matrix when you are considering more than one function. In fact, now that we know V_y we can go on to consider other functions based on y_1 and y_2. For example, consider the function $z = y_1 + y_2$. Straightaway we can see that z must equal 1 from the definitions of y_1 and y_2. But what is the error on z?

Our new function has two variables, which according to equation C.6 means that the error in z can be determined from $V_z = D^T V_y D$, where in this case $D = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Thus:

$$\begin{aligned} V_z &= \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 0.0052 & -0.0052 \\ -0.0052 & 0.0052 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &= 0 \end{aligned}$$

Thus $z = 1 \pm 0$, which is correct because by definition the sum of y_1 and y_2 must always be equal to one and can thus have no uncertainty.

Note the effect if we had ignored the covariance between y_1 and y_2 we would get:

$$y_1 = 0.4 \pm 0.072 \qquad \text{and} \qquad y_2 = 0.6 \pm 0.072$$

Hence $z = 1$ and $\sigma_z^2 = (0.072)^2 + (0.072)^2 \neq 0$ which is clearly incorrect.

Thankfully, the covariance term keeps track of the fact that y_1 and y_2 are not independent and has prevented us from making this blunder!

Exercise C.3

To emphasize this point, imagine $y_1 = x$ and $y_2 = x^2$. y_1 and y_2 are clearly correlated since they both depend on x. What is the uncertainty in $z = y_2/y_1$? You can see by inspection that the answer should be $\sigma_z^2 = \sigma_x^2$, whereas ignoring covariance and following the "normal" rules of error propagation you would get $\sigma_z^2 = 5\sigma_x^2$, which is incorrect.

Now that you are (hopefully) all convinced that it is important to calculate the covariance terms, we return to the problem we have been working on over the last two weeks – least squares fitting.

C.4 Covariance and Least Squares Fitting

It turns out that one place where the derived quantities are likely to be correlated is in least squares fitting. In fact, it is highly *unlikely* that the gradient (m) and intercept (c) determined from a straight line fit are *uncorrelated*. Hence we need to calculate $\text{cov}(m, c)$ as well as σ_m^2 and σ_c^2 in order to properly calculate the error on quantities that depend on both the gradient and intercept.

The equations for deriving the parameters for a straight line least squares fit were given in Session B and can be written in matrix form (reverting to the nomenclature of Session B that p_0 is the

intercept and p_1 the gradient):

$$\begin{pmatrix} [\omega x^2] & [\omega x^1] \\ [\omega x^1] & [\omega x^0] \end{pmatrix} \begin{pmatrix} p_1 \\ p_0 \end{pmatrix} = \begin{pmatrix} [\omega x^1 y] \\ [\omega x^0 y] \end{pmatrix} \tag{C.8}$$

Writing this in brief form:

$$A \begin{pmatrix} p_1 \\ p_0 \end{pmatrix} = \begin{pmatrix} [\omega x^1 y] \\ [\omega x^0 y] \end{pmatrix} \tag{C.9}$$

where

$$A = \begin{pmatrix} [\omega x^2] & [\omega x^1] \\ [\omega x^1] & [\omega x^0] \end{pmatrix} \tag{C.10}$$

we can see that the solution is (using matrix formalism):

$$\begin{pmatrix} p_1 \\ p_0 \end{pmatrix} = A^{-1} \begin{pmatrix} [\omega x^1 y] \\ [\omega x^0 y] \end{pmatrix} \tag{C.11}$$

It turns out that in determining the A^{-1} matrix in order to derive the solutions for p_0 and p_1 we have done most of the work required to find the full covariance matrix, $V_{p_1 p_0}$, for the gradient and intercept. However, there are some subtleties to consider, which have to do with how well the errors on the data (and hence the weights) are determined.

Some texts (e.g. Smith – see further reading at the end) say that the covariance matrix is given simply by A^{-1}:

$$V_{p_1 p_0} = \begin{pmatrix} \sigma_{p_1}^2 & \mathrm{cov}(p_1, p_0) \\ \mathrm{cov}(p_1, p_0) & \sigma_{p_0}^2 \end{pmatrix} = A^{-1} \tag{C.12}$$

Others (e.g. Martin) quote:

$$V_{p_1 p_0} = \frac{\chi^2}{dof} A^{-1} \tag{C.13}$$

That is A^{-1} multiplied by *chi-squared* divided by the number of degrees of freedom. The number of degrees of freedom is equal to the number of data points minus the number of free parameters. For a straight line fit the number of free parameters is two (p_1 and p_0). Hence the chi-squared per degree of freedom is given by:

$$\frac{\chi^2}{dof} = \frac{\sum_{i=1}^{n} \omega_i \left(y_i - [p_1 x_i + p_0]\right)^2}{n-2} \tag{C.14}$$

So, which formula is correct? In an ideal world it ought not to matter, since if your errors are well known the χ^2/dof should be close to unity anyway. However, real-life situations are rarely ideal, so some care is needed.

Let us consider two examples:

In cases where you have input data for which, (a) the errors are well known, (b) there is no reason to believe that the resulting errors on y are either under or over-estimates and (c) the order of the polynomial you are trying to fit is known (perhaps on theoretical grounds), then it should be fine to assum that $V_{p_1 p_0} = A^{-1}$. However, you should still look at the χ^2/dof to see if it is significantly different from unity. If it is, then you must go back to the data to look for values that may be

outliers – these are points that lie far away from the predicted curve. Is there some reason that these outliers exist? Can you conceive of grounds for eliminating them? Alternatively, you must question the errors on the input data. Can you be sure they are correct?

Alternatively, you may have data for which you are unsure of the errors or have no grounds for seeking any particular order of polynomial. You might simply want some convenient way of describing the data distribution. In this case the χ^2/dof can be different from unity because the input errors are wrong, or the data does not fit this polynomial or you have outliers that ought to be discarded. Often you can do little about the input errors, unless you have the opportunity to remake the measurement, so the only thing to do is to rescale all the errors to give a reasonable final χ^2/dof. This is what you achieve by using equation C.13. It is the same as scaling all the input standard deviations by $(\chi^2/dof)^{1/2}$ and then using equation C.12 on this new set of data. If you do not know which polynomial to use, you have no reason to expect that the χ^2/dof should be equal to unity. Faced with an uncertain hypothesis you would generally choose a function that results in χ^2/dof closest to unity. However, your "best" function may still give χ^2/dof a long way from unity. In this case, the "best" estimate for the error on the parameters is given by equation C.13.

POLYFIT shows both forms for the covariance matrix. In the absence of other arguments, I would suggest that you always use equation C.13.

The least squares programs that you meet in the laboratory often only print the errors on the gradient and intercept. These are simply the square roots of the diagonal terms in the covariance matrix, $V_{p_1 p_0}$. If you only need the gradient or intercept then these are the appropriate errors to use. However, it is important to realise that if you combine the gradient and intercept in order to derive a further quantity you will only get the correct error on that quantity if you use the full covariance matrix for p_1 and p_0, and use equation C.2 (or in its matrix form equation C.6) above. For example, consider the calculation of the intercept on the x-axis ($= -p_0/p_1$). The error on this must be calculated using equation C.2 and the complete covariance matrix as it requires the use of both p_1 and p_0. The final exercise in this session is designed to reinforce these ideas.

Exercise C.4

In a compound pendulum experiment in a teaching laboratory one can derive a value for the acceleration due to gravity, g, from the gradient (p_1) of a least square fit of t^2h versus h^2, where t is the period of the pendulum and h is its length. The relationship between g and p_1 is $g = 4\pi^2/p_1$. The radius of gyration, k, is also derived from $(\text{intercept}/\text{gradient})^{1/2}$. Thus while we can ignore $\text{cov}(p_1, p_0)$ in calculating the error on g (because g is a function of p_1 only), we do need to use $\text{cov}(p_1, p_0)$ when calculating the error on k.

Thus, using data from a student experiment, we obtain:

Supplied Data: h^2	Supplied Data: ht^2
0.170982	1.0283
0.177662	1.0509
0.126025	0.8430
0.127092	0.8499
0.093025	0.7093
0.091506	0.7063
0.060270	0.5785
0.057600	0.5670
0.042230	0.5065
0.023562	0.4296
0.013110	0.3885
0.041820	0.5041
0.022500	0.4234
0.013572	0.4009

Use Excel (or POLYFIT) to show that the result is:

$$\begin{aligned}\text{Gradient} &= 4.0276\\ \text{Intercept} &= 0.33665\end{aligned}$$

The covariance matrix (equation C.13) for this pair of values is:

$$\begin{matrix} 2.73032 \times 10^{-4} & -2.06911 \times 10^{-5} \\ -2.06911 \times 10^{-5} & 2.37461 \times 10^{-6} \end{matrix}$$

Using the relationship that the gradient is equal to $4\pi^2/g = 4.0276$, show that the numerical value of $g = 9.80 \pm 0.04$ ms^{-2}.
Also, given that $k = (p_0/p_1)^{1/2}$, show that $k = 0.2891 \pm 0.0012$.
Note that you will need to use $\text{cov}(p_1, p_0)$ to calculate the uncertainty.
(Notice that if we had ignored the covariance term we would have obtained $\sigma_k = 0.0009$. In this case the difference is relatively small, but it serves to illustrate the point).

C.5 Alternative form for the covariance matrix

The previous discussion has assumed that we can calculate the elements of the matrix A, and hence determine the covariance matrix from A^{-1} multiplied by χ^2/dof. However, for cases where the function to be fitted is not a polynomial, we cannot calculate A analytically (see Session B.2). In these cases, we must find the values of the parameters using iterative methods, but what about the errors on those parameters?

It turns out that the matrix A can also be written as:

$$A = \begin{pmatrix} \frac{1}{2}\frac{\partial^2\chi^2}{\partial p_1^2} & \frac{1}{2}\frac{\partial^2\chi^2}{\partial p_1 \partial p_0} \\ \frac{1}{2}\frac{\partial^2\chi^2}{\partial p_1 \partial p_0} & \frac{1}{2}\frac{\partial^2\chi^2}{\partial p_0^2} \end{pmatrix} \tag{C.15}$$

You can verify this for the straight line fitting case by comparing what you get using the above expression for A with equation C.10. How this form can be extended for more parameters p_2, p_3 etc. should be obvious.

The advantage of this form is that the required derivatives can be calculated numerically (see Session F), so there is no need to know an analytic form for A. The disadvantage is that the numerical values of the derivatives will be limited in their accuracy according to the approximations made in calculating them. We won't discuss this further, as it leads to concerns over what the error on the error should be...

Summary

Finally, *always* assume that errors on parameters are correlated unless you know for certain that they are not. That is, always assume that you need to use equation C.2 (or C.6). In most real-life situations you can rarely get away with using equation C.1.

Also, only use equation C.15 when an analytical form for A is impossible to determine, as the approximations made to determine the derivatives can reduce the accuracy.

Further reading

Bevington P. R. "Data Reduction and Error Analysis for the Physical Sciences.", McGraw Hill (1969)
Hildebrand F. B. "Introduction to Numerical Analysis", Courier Corporation (1987)
Martin B. R. "Statistics for Physicists.", Academic Press (1971)
Smith D. L. "Probability, Statistics and Data Uncertainties in Nuclear Science and Technology", OECD (1991)
Owen J.G., Weaver D. R. and Walker J. *Nucl. Instrum. and Methods* **188**, 579–593 (1981) (some research based examples of use of covariances).